SUPPORT VECTOR MACHINE (SVM) PROJECT REPORT

Introduction

The Support Vector Machine (SVM) project focuses on building a classification model to distinguish between two classes in a dataset. SVM is a powerful supervised learning algorithm used for both classification and regression tasks. The main goal of this project is to classify whether a user will purchase a product on a social networking site based on age and estimated salary.

Dataset and Algorithms

For this project, I worked with a dataset named "Social_Network_Ads.csv," containing information about social network ads. The dataset consists of features such as age and estimated salary, and the target variable is whether the user purchased the advertised product (binary classification).

Data Preprocessing

I began by importing necessary libraries for data manipulation, visualization, and modeling. Next, I loaded the dataset and split it into features (X) and the target variable (y). The dataset was then divided into training and test sets using a test size of 25% and a random state of 0 for reproducibility.

To ensure accurate modeling, I applied feature scaling using StandardScaler, standardizing the data to create a normal distribution of the features.

Model Training and Evaluation

I chose the Support Vector Machine (SVM) algorithm with a linear kernel to build the classification model. SVM is suitable for binary classification problems and can handle both linearly and non-linearly separable data.

After training the SVM model on the training set, I used it to predict the class labels for the test set. I evaluated the model's performance using metrics such as the confusion matrix, which provides insights into the number of correct and incorrect predictions, and the accuracy score, which measures the fraction of correct predictions made by the model.

Result Visualization

To gain a visual understanding of the model's performance, I created scatter plots for both the training and test sets. These plots show the data points colored according to their actual class labels, with "Not Purchased" represented in salmon and "Purchased" in dodger blue. This visualization provides a clear view of how well the model distinguishes between the two classes.

Conclusion

The SVM project showcases the effectiveness of the Support Vector Machine algorithm in handling binary classification tasks. By accurately classifying social network ad data into "Purchased" and "Not Purchased" categories, the SVM model can be useful for predicting user behavior based on age and estimated salary.

A real-life usage instance of this project could be in the field of marketing, where companies use this type of algorithm to target potential customers based on their demographic data and previous buying behavior, e.g. a company selling a high-end product might use this algorithm to identify potential customers with high incomes and a history of purchasing similar products.

Note: The code provided in the script demonstrates the implementation and evaluation of the SVM model for this specific project. Further improvements, optimization, and integration with other applications can be explored to enhance the model's usability and impact.

Thank you for taking the time to explore the SVM project! If you have any questions or feedback, feel free to reach out to me at osuolalefolarin@gmail.com

GitHub Repository: https://github.com/Folarinosuolale/Data-Science-Machine-Learning